Molecular chaperones: **Clasping the prize**

Mary-Jane Gething

The three-dimensional structure of the substrate-binding domain of DnaK, a bacterial Hsp70, shows how such molecular chaperones can be so promiscuous in recognizing different proteins, yet so accurate in discriminating between unfolded and folded forms of their polypeptide substrates.

Address: Department of Biochemistry and Molecular Biology, University of Melbourne, Parkville, Victoria 3052, Australia.

Current Biology 1996, Vol 6 No 12:1573–1576

Heat-shock proteins are induced in cells subject to thermal and other stresses, and many of these are now known to be molecular chaperones. Perhaps best known is the heat-shock protein 70 (Hsp70) class, a family of highly conserved and ubiquitous proteins found in all organisms, from bacteria to man, and in all subcellular compartments of eukaryotic cells [1,2]. Hsp70s perform a multitude of functions in cells, requiring considerable versatility: they must recognize and bind to a wide variety of target proteins that share no obvious sequence similarity, but must also distinguish native from unfolded proteins. These functions include participating in protein synthesis, stabilizing newly synthesized or unfolding polypeptides, facilitating translocation of nascent chains across membranes and into the nucleus, mediating the assembly or disassembly of multimeric protein complexes and targeting proteins for degradation within lysosomes [1,3,4]. The *in vivo* substrates of Hsp70s are partially folded polypeptides; *in vitro* studies with individual family members demonstrate that peptides as short as 5–8 residues bind stably to Hsp70s and stimulate their ATPase activity [5–7].

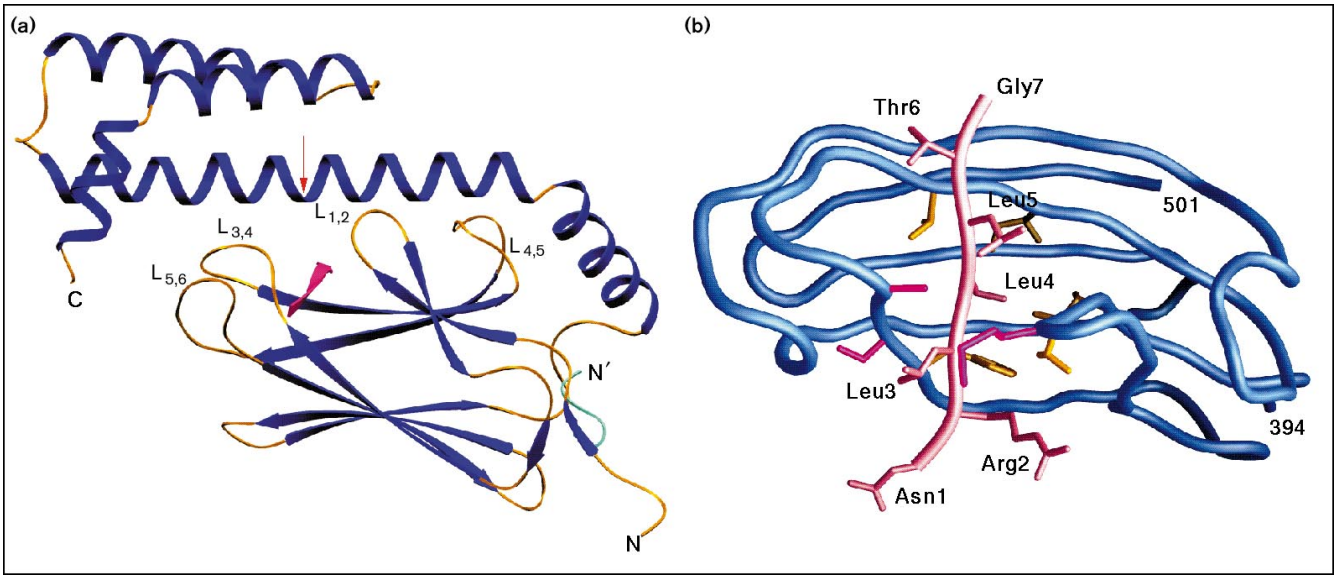
Hsp70s contain two major domains, an amino-terminal domain that contains the ATPase catalytic site and a carboxy-terminal substrate-binding domain. These domains communicate to regulate the affinity and duration of polypeptide binding [8]. Thus, the affinity of the carboxy-terminal domain for polypeptides depends on whether the nucleotide-binding site in the amino-terminal domain is occupied by ATP or ADP, while binding of a polypeptide to the carboxy-terminal domain increases the rate of ATP hydrolysis by the amino-terminal ATPase domain. The elucidation six years ago of the three-dimensional structure of the isolated ATPase domain of the constitutively-expressed bovine cytosolic Hsc70 [9] was an important milestone in studies on Hsp70 chaperones. The recent determination of an X-ray crystallographic structure of a carboxy-terminal fragment of DnaK, an Hsp70 from *Escherichia coli*, with bound peptide [10] ushers in a new era in Hsp70 research.

Perhaps the most intriguing question about Hsp70s has been how they can be so accurate in discriminating between unfolded and folded forms of their polypeptide substrates, and yet so promiscuous in recognizing an extremely wide variety of different proteins. Half of the answer to this paradox was supplied by comparative analyses of the sequences of large collections of short peptides that bind to DnaK ([7] and B. Bukau, personal communication), Hsc70 [11] or BiP (the major Hsp70 in the endoplasmic reticulum) [5,6]. Although individual Hsp70s show some variation in their peptide-recognition patterns ([12,13] and see below), they all share a predilection for short sequences (~5–7 amino acids) that are enriched in hydrophobic residues. Thus, the Hsp70 chaperones bind sequences that are likely to be exposed only by partially or completely unfolded polypeptides; such sequences would normally be buried in the interior of fully folded proteins, the surfaces of which are normally coated with polar and charged residues. The peptides recognized by Hsp70s do not share any specific sequence motif, but rather display great diversity in the identity and positioning of the preferred hydrophobic residues. This is consistent with the observed promiscuity of binding.

The second half of the answer — how the chaperones bind such diverse sequences — has now been largely provided by the determination of the structure of the substrate-binding domain of DnaK with a bound heptapeptide [10]. The X-ray crystallographic analysis, carried out to 2.0 Å resolution, revealed a completely novel fold, compatible with secondary-structure assignments obtained for the equivalent domain of bovine Hsc70 by nuclear magnetic resonance (NMR) spectroscopy [14], but disproving earlier predictions that the structure might resemble the peptide-binding platform of major histocompatibility complex (MHC) class I molecules.

The substrate-binding domain of DnaK consists of a compact β sandwich followed by helical elements (Fig. 1a). The β sandwich is arranged in two sheets with four antiparallel β strands in each, and the peptide is bound in an extended conformation in a channel formed by loops that connect strands of the sandwich. The particular topology of this β sandwich has not been observed previously in other protein structures. The five helical elements are not directly involved in binding to the peptide, but rather stabilize the β domain and act as a lid to encapsulate the peptide. Interestingly, this five-helix unit is similar in

Figure 1



Three dimensional structure of the substrate-binding domain of DnaK [10]. **(a)** Ribbon diagram of a Cα trace of the structure of the DnaK substrate-binding domain, with peptide (pink) bound in a channel formed by loops off the β sandwich and capped by an α-helical domain. Loops $L_{1,2}$ and $L_{3,4}$ directly form the channel for substrate binding, while loops $L_{4,5}$ and $L_{5,6}$ stabilize the channel by buttressing loops $L_{1,2}$ and $L_{3,4}$, respectively, through main-chain and side-chain hydrogen bonds and hydrophobic interactions. The red arrow shows the point at which the long αB-helix kinks upward in the type 2 crystal structure. The amino terminus of the substrate-binding domain, which in the intact DnaK molecule would be connected to the ATPase domain, lies almost at the opposite end of the structure from the substrate channel and the α-helical lid. The amino-terminal residues can take up two alternative conformations, the second (N′) being shown in blue-green. **(b)** Binding of peptide Asn–Arg–Leu–Leu–Leu–Thr–Gly to the β subdomain of DnaK. The protein and the peptide backbones are shown in blue and pink, respectively. The side chains of the peptides are labelled, and the DnaK residues interacting with the side chain of Leu 4 of the peptide are coloured yellow, and those interacting with Leu 3 and Leu 5 of the peptide are coloured magenta. (Images generously prepared and provided by Wayne Hendrickson and Xiaotian Zhu.)

topology and general structure to a portion of a helical domain which entraps the bound guanine nucleotide in the α subunits of trimeric G proteins [15,16].

The bound peptide, which has the sequence Asn–Arg–Leu–Leu–Leu–Thr–Gly, was originally identified in a bacteriophage display screen for peptides that bind DnaK with high affinity [7]. Binding interactions between this peptide and DnaK are centred and concentrated on Leu 4, which is completely buried in a deep pocket, and involve almost exclusively the middle five residues of the heptapeptide (Fig. 1b) [10]. Leu 3 is also highly buried, and Leu 3 and Leu4 together contribute most of the side-chain contacts with DnaK. Arg 2, Leu 5 and Thr 6 are moderately buried and contribute the remainder of the side-chain contacts. The terminal residues, Asn 1 and Gly 7, make no significant side-chain contacts with DnaK, although the question of whether a residue other than glycine at position 7 could provide stabilizing side-chain interactions remains open.

The side-chain van der Waals contacts between the peptide and DnaK are dominated by hydrophobic interactions, consistent with the predictions from peptide-binding studies. The peptide main-chain contributes seven hydrogen bonds, five of which are made with DnaK backbone groups. The pocket that binds Leu 4, designated the central site '0' by Zhu *et al.* [10], appears to be the crucial peptide-binding determinant. Contact analysis indicates that leucine is the ideal occupant for this site, but methionine and isoleucine could also fit, albeit very snugly [10]. Adjustments of the protein or peptide backbone might allow the pocket to accept phenylalanine, but it probably could not accommodate tyrosine or tryptophan. Smaller side chains would fit, paying an energetic penalty for the residual cavity, which in the cases of threonine or serine might be compensated for by formation of a hydrogen bond to the side-chain oxygen atom of Thr 437 in the β4 strand of the sandwich.

To what extent the peptide-binding channel might be able to mould itself to a wider range of residues in the anchoring position, as the loops close to form the 0 site, will only be revealed when structures become available for DnaK substrate-binding domains with other peptides bound. Geometric constraints do not appear to provide significant restrictions on the identity of residues at positions in the peptide other than the anchor residue at site 0, although the observed peptide ψ angles appear to exclude

proline. A general hydrophobic character at sites –1 and +1 is generated by a hydrophobic arch over the binding channel that is constructed by contacts between the two loops that form the channel (see Fig. 1b).

Sites outside the central triad are quite open, and the electrostatic potential of the surface of DnaK at these positions is largely negative, accounting for the observed exclusion of acidic residues and prevalence of basic residues at the amino and carboxyl termini of hexapeptides and heptapeptides that bind DnaK [7]. Bernd Bukau and colleagues (personal communication) have recently identified DnaK binding sites within 4360 overlapping 13-residue peptides that scan 37 protein sequences. The binding motif that emerged from this study consists of a 4–5 residue hydrophobic core, particularly enriched in leucine, flanked by short basic segments. This motif is consistent with earlier observations of Gragerov *et al.* [7] using peptides displayed by bacteriophages, and is completely compatible with the structure of the DnaK substrate-binding site [10].

Although we now have a very clear picture of how at least one peptide fits in the substrate-binding channel of an Hsp70, the question remains as to how the substrate enters and exits the binding site. As Zhu *et al.* [10] point out, it would be extremely difficult for even a short peptide to thread its way into or out of the DnaK binding channel. For a natural substrate, in which the binding site is part of a continuous (and often partly folded) polypeptide chain, a threading mechanism of entry or exit would be impossible. It has long been recognized from protease-sensitivity and spectroscopic studies (reviewed in [17]) that large conformational changes take place in the substrate-binding domain of Hsp70 proteins, potentiated first when ATP binds to the amino-terminal domain, promoting release of bound polypeptide, and then when ADP is generated following hydrolysis of the ATP, which induces a high-affinity state for substrate binding.

Although an X-ray structure can present only a stationary snapshot of one stage of the 'Hsp70 reaction cycle' [4] of substrate binding and release, analysis of two different crystal forms of the complex provides important clues as to the mechanism of this process [10]. One such clue is that the contacts between the α-helical lid that caps the binding channel and the loops that form the channel involve only a relatively small buried surface area. Another is that the structures of the complex in the two crystal forms differ only in that, in the 'type 2' form, these contacts are disrupted and the 'lid' is partially lifted by a hinge-like transformation around a kink half way along the capping helix (Fig. 1a). The peptide remains bound in the type 2 structure, but is less well ordered, and Zhu *et al.* [10] consider the type 2 conformation as being incipient to the full-scale change required for peptide exchange. It thus appears that, on binding of ATP to the amino-terminal domain, the helical domain hinges up and away, uncapping the peptide channel.

Even once this uncapping is achieved, substantial movement of the channel-forming loops must occur before a bound polypeptide can be released from their clasp, or a new polypeptide can enter the binding site. Binding of the new substrate would itself cause a conformational change that, once propagated to the amino-terminal domain, would stimulate ATP hydrolysis — and the presence of ADP in the nucleotide-binding site would in turn cause a conformational change in the ATPase domain that, when propagated back to the carboxy-terminal domain, would stabilize binding of the substrate. Replacement of the bound ADP by ATP, a rate-limiting step for DnaK that is accelerated by the co-chaperone GrpE [4], would restart the cycle of polypeptide release and binding. The second co-chaperone, DnaJ, which potentiates the substrate-induced stimulation of ATP hydrolysis [4], is thought to bind to the region of DnaK corresponding to the α-helical domain [18]. Zhu *et al.* [10] suggest that binding of peptide or DnaJ may stabilize the closed state of the α-helical lid, thus favouring the ADP-bound conformation and consequently stimulating ATP hydrolysis.

Although no structure is yet available for an intact Hsp70 molecule, the structure of the isolated substrate-binding domain provides significant insight into how the amino-terminal and carboxy-terminal domains interact to facilitate coupling of the conformational changes that occur upon the binding of polypeptide and nucleotides. The amino terminus of the substrate-binding domain, which in the intact molecule would be connected to the ATPase domain, lies almost at the opposite end of the structure from the substrate channel and the α-helical lid [10] (see Fig. 1a). Thus, conformational changes in the ATPase domain cannot directly affect the structure of the lid. Inspection of two alternative dispositions of the amino-terminal residues of the substrate-binding domain, in the type 1 crystal, indicate that these residues can be either extended or folded into a hydrophobic pocket near the junction between the carboxyl terminus of the β sandwich domain and the beginning of the α-helical domain (Fig. 1a). Zhu *et al.* [10] surmise that conformational changes in the ATPase domain could displace the amino-terminal sequence from the hydrophobic pocket, causing changes in the positioning of the α-helical domain relative to the β sandwich and affecting the equilibria between the open and closed states of the lid.

What lessons does the DnaK structure provide for other members of the Hsp70 family? Zhu *et al.* [10] note that the pattern of sequence conservation within the Hsp70 family is such that the backbone conformation in the β subdomain would be virtually identical in all Hsp70s. Furthermore, all of the backbone hydrogen bonds and one of

the two side-chain hydrogen bonds between DnaK and the peptide backbone should be maintained in all members of the family. Apart from the central pocket, which is very highly conserved, residues that in DnaK contact the side-chains of the peptide can vary in the different family members, and the electrostatic potential of the surfaces surrounding the substrate-binding channel is also likely to vary in different Hsp70 proteins. Such similarities and differences between the peptide-binding channels of Hsp70s could account for the overlapping but distinct peptide-binding specificities exhibited by DnaK, Hsc70 and BiP [12,13].

While DnaK, Hsc70 and BiP generally show similar high affinities for highly hydrophobic peptides, significant differences in their binding affinities are quite frequently observed. An example that can be readily explained using the new structural information is the differential interaction of the chaperones with the heptapeptide Lys–Lys–Leu–Met–Phe–Lys–Thr. This peptide binds with high affinity to DnaK, with moderate affinity to Hsc70 and with negligible affinity to BiP [12]. The relative affinities of these proteins are consistent with the observation of Zhu *et al*. [10] that the surface of the peptide-binding channel of DnaK (and probably Hsc70) is electronegative at the amino-terminal side of the peptide, whereas in BiP this surface is likely to be slightly electropositive (possibly resulting in charge repulsion of the amino-terminal lysine residues). In support of this explanation, substitution of the lysine at position 2 in the heptapeptide by tyrosine essentially abolished the specificity difference between the three Hsp70 proteins by increasing the affinities of DnaK and Hsc70 5-fold and 20-fold, respectively, and that of BiP by greater than two orders of magnitude [12].

Further evidence that the peptide-binding preferences of Hsp70 proteins vary in detail comes from the algorithms that have been used with success to predict sequences likely to bind to BiP [6,19] or DnaK (B. Bukau, personal communication). Each algorithm was developed by comparing the relative occurrences of amino acids in ~100 binding and non-binding sequences. In the case of BiP, the algorithm describes a heptameric motif that is enriched in hydrophobic and aromatic residues, particularly in alternating positions, whereas for DnaK the algorithm describes a five-residue hydrophobic core flanked by basic residues. Interestingly, of nine BiP-binding sequences identified within an immunoglobulin heavy chain using the BiP 'Score' algorithm [19], only two bind appreciably to DnaK (M. Beasley and M-J.G., unpublished observation). Significantly, the DnaK Score algorithm accurately discriminates these two likely binding sequences from the others, indicating that the algorithms do reflect the distinct binding specificities of the two chaperones. How well the motif defined for BiP, which suggests BiP has four binding pockets for bulky hydrophobic residues, describes the actual binding site on the BiP molecule, and how differential binding specificities of DnaK and BiP (and other Hsp70 proteins) are reflected in the detailed topographies of their substrate-binding sites, must await the elucidation of additional structures for BiP and other members of the Hsp70 family.

References

1. Gething MJ, Sambrook JF: Protein folding in the cell. *Nature* 1992, 355:33–45.
2. Boorstein WR, Ziegelhoffer T, Craig EA: Molecular evolution of the HSP70 multigene family. *J Mol Evol* 1994, 38:1–17.
3. Becker J, Craig EA: Heat-shock proteins as molecular chaperones. *Eur J Biochem* 1994, 219:11–23.
4. Hartl FU: Molecular chaperones in cellular protein folding. *Nature* 1996, 381:571–580.
5. Flynn GC, Pohl J, Flocco MT, Rothman JE: Peptide-binding specificity of the molecular chaperone BiP. *Nature* 1991, 353:726–730.
6. Blond-Elguindi S, Cwirla SE, Dower WJ, Lipshutz RJ, Sprang SR, Sambrook JF, Gething MJ: Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BiP. *Cell* 1993, 75:717–728.
7. Gragerov A, Zeng L, Zhao X, Burkholder W, Gottesman ME: Specificity of DnaK–peptide binding. *J Mol Biol* 1994, 235:848–854.
8. McKay DB: Structure and mechanism of 70-kDa heat-shock-related proteins. *Adv Prot Chem* 1993, 44:67–80.
9. Flaherty KM, DeLuca-Flaherty C, McKay DB: Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein. *Nature* 1990, 346:623–628.
10. Zhu X, Zhao X, Burkholder WF, Gragerov A, Ogata CM, Gottesman ME, Hendrickson WA: Structural analysis of substrate binding by the molecular chaperone DnaK. *Science* 1996, 272:1606–1614.
11. Takenaka IM, Leung S-M, McAndrew SJ, Brown JP, Hightower LE: Hsc70-binding peptides selected from a phage display peptide library that resemble organellar targeting sequences. *J Biol Chem* 1995, 270:19839–19844.
12. Fourie AM, Sambrook JF, Gething M-JH: Common and divergent peptide-binding specificities of hsp70 molecular chaperones. *J Biol Chem* 1994, 269:30470–30478.
13. Gragerov A, Gottesman ME: Different peptide binding specificities of hsp70 family members. *J Mol Biol* 1994, 241:133–135.
14. Morshauser RC, Wang H, Flynn GC, Zuiderweg ERP: The peptide-binding domain of the chaperone protein Hsc70 has an unusual secondary structure topology. *Biochemistry* 1995, 34:6261–6266.
15. Noel JP, Hamm HE, Sigler PB: The 2.2 Å crystal structure of transducin-α complexed with GTPγS. *Nature* 1993, 366:654–663.
16. Coleman D, Berghuis AM, Lee E, Linder ME, Gilman AG, Sprang SR: Structures of active conformations of $G_{i\alpha 1}$ and the mechanism of GTP hydrolysis. *Science* 1994, 265:1405–1412.
17. Fung KL, Hilgenberg L, Wang N, Chirico WJ: Conformations of the nucleotide and polypeptide binding domains of cytosolic Hsp70 molecular chaperones are coupled. *J Biol Chem* 1996, 271:21559–21565.
18. Wawrzynów A, Zylicz M: Divergent effects of ATP on the binding of the DnaK and DnaJ chaperones to each other, or to their various native and denatured protein substrates. *J Biol Chem* 1995, 270:19300–19306.
19. Knarr G, Gething M-J, Modrow S, Buchner J: BiP-binding sequences in antibodies. *J Biol Chem* 1995, 270:27589–27594.